European Journal of Operational Research 277 (2019) 32–41

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Continuous Optimization

A multilevel analysis of the Lasserre hierarchy

Juan S. Campos*, Ruth Misener, Panos Parpas

Department of Computing, Imperial College London, London SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 28 June 2018
Accepted 6 February 2019
Available online 12 February 2019

Keywords:
Global optimization
Conic programming and interior point methods
Semidefinite programming
Polynomial optimization

ABSTRACT

This paper analyzes the relation between different orders of the Lasserre hierarchy for polynomial optimization (POP). Although for some cases solving the semidefinite programming relaxation corresponding to the first order of the hierarchy is enough to solve the underlying POP, other problems require sequentially solving the second or higher orders until a solution is found. For these cases, and assuming that the lower order semidefinite programming relaxation has been solved, we develop prolongation operators that exploit the solutions already calculated to find initial approximations for the solution of the higher order relaxation. We can prove feasibility in the higher order of the hierarchy of the points obtained using the operators, as well as convergence to the optimal as the relaxation order increases. Furthermore, the operators are simple and inexpensive for problems where the projection over the feasible set is "easy" to calculate (for example integer {0, 1} and $\{-1, 1\}$ POPs). Our numerical experiments show that it is possible to extract useful information for real applications using the prolongation operators. In particular, we illustrate how the operators can be used to increase the efficiency of an infeasible interior point method by using them as an initial point. We use this technique to solve quadratic integer {0, 1} problems, as well as MAX-CUT and integer partition problems.

1. Introduction

Tight convex relaxations are the most valuable tool in the optimizer's toolbox for the approximate solution of NP-hard problems (Boukouvala, Misener, & Floudas, 2016). The Lasserre and other related hierarchies are one such incredibly powerful relaxation for polynomial optimization problems (POPs). Unfortunately, these hierarchies require solving Semidefinite Programming (SDP) problems that grow exponentially with the relaxation order, limiting the use of interior point methods (IPM). To address this issue, a great deal of research has gone into exploiting special mathematical structure (de Klerk, 2010) and developing different hierarchies (Ahmadi & Majumdar, 2017; Lasserre, Toh, & Yang, 2017; Weisser, Lasserre, & Toh, 2017). The sparse relaxations proposed in Waki, Kim, Kojima, and Muramatsu (2006) and further analyzed in Lasserre (2006), enabled an order of magnitude improvement in terms of the dimensionality of problems that can be solved with sum of squares (SOS) relaxations. Specialized algorithms such as the low rank approximations developed in Burer and Monteiro (2003) and the semi-smooth CG and alternating direction augmented Lagrangian methods in Yang, Sun, and Toh (2015) and Wen, Goldfarb, and Yin (2010), respectively, have also helped address the computational issues associated with solving large-scale problems.

Despite this progress, the issue of SDP relaxations whose size increases exponentially with the order of the relaxation persists. We take a step towards addressing this issue by developing linear operators called prolongation operators for the Lasserre hierarchy. These operators transfer information from a hierarchy of order w to a hierarchy of order $w+1$. The prolongation operators allow us to approximate both the primal and dual solutions of the relaxation of order $w+1$, by only using information from the order w relaxation. A crucial property of the proposed operators is that, concerning computational effort, they are virtually free and are easy to implement. Campos and Parpas (2018) develop prolongation operators that are used to transfer information between different optimization problems through a single Lasserre hierarchy. Besides, their computation requires no parameters or any additional assumptions. The links between the solutions of relaxations at different hierarchies are studied here for the first time. We develop the proposed operators using the original Lasserre hierarchy, but the results can easily be extended to the sparse hierarchy in Waki et al. (2006). We anticipate the proposed approach to be applicable to the study of other SOS relaxations such as BSOS (Lasserre et al., 2017), but this work focuses on the most widely used hierarchy.

Although our numerical results show that the operators can be used to construct useful initial points for warm start strategies, it

* Corresponding author.
E-mail addresses: jsc12@ic.ac.uk (J.S. Campos), r.misener@ic.ac.uk (R. Misener), panos.parpas@ic.ac.uk (P. Parpas).

https://doi.org/10.1016/j.ejor.2019.02.016
0377-2217/

is important to remark that our results are still bounded by the limits of the SDP relaxations for POP: first, the number of variables in the polynomial space that we are able to handle is limited, and second, our method relies on an SDP algorithm that can take advantage of good initial points. In this paper we use interior point methods as our base algorithm. Unfortunately, exploiting initial points is a challenging and open problem for interior point algorithms, and this issue sets an upper bound on the performance of the prolongation operators developed in this paper.

We consider the following constrained polynomial optimization problem (POP):

$$\begin{aligned} p^\star := \min_{\mathbf{x}\in\mathbb{R}^n} & f(\mathbf{x}) \\ s.t.\ & h_i(\mathbf{x}) \geq 0,\ i = 1,2,\ldots,m, \end{aligned} \tag{1}$$

where f and h_i $(i = 1,2,\ldots,m)$, are n-dimensional polynomial functions with degrees d and $d_1, d_2, \ldots, d_m$, respectively. In addition to the usual (and generally non-restrictive) assumptions for the convergence of the Lasserre hierarchy to polynomial optimization problems, we make the following assumption.

Assumption 1.1. The feasible set $K = \{\mathbf{x}\in\mathbb{R}^n : h_i(\mathbf{x}) \geq 0,\ i = 1,2,\ldots,m\}$ is compact and such that the projection of any $\mathbf{x}\in\mathbb{R}^n$ onto the set K is tractable.

Assumption 1.1 is not strictly necessary from a theoretical point of view. But computing the prolongation operators requires a projection into the feasible set, and therefore Assumption 1 is needed from a practical point of view. We note that many open problems satisfy Assumption 1, including MAX-CUT (Caprara, 2008), partitioning (van Dam & Sotirov, 2015), and generic polynomial 0/1 programs (Lasserre, 2016).

The principal theoretical contribution of this work is to provide insight into the relationship between different relaxation orders. In particular, Section 3 establishes connections between the input data of relaxations of different orders. We then develop our operators for both the primal and dual variables and establish the feasibility characteristics of the prolongated variables. From a practical point of view, the proposed operators can be used to construct an initial point for an optimization algorithm. In Section 5, we show that the calculation of initial points using our operators can improve the solution times of interior point methods when they are used in combination with a warm start strategy.

2. Notation

Given a real-valued polynomial function $f : \mathbb{R}^n \to \mathbb{R}$ of degree d, let the monomial $x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}$ be denoted by $\mathbf{x}^{\boldsymbol{\alpha}}$ and its coefficient by $b_{\boldsymbol{\alpha}}$, where $\boldsymbol{\alpha}\in\mathbb{N}^n$. If $\Gamma_d^n = \{\boldsymbol{\alpha}\in\mathbb{N}^n : \sum_i \alpha_i \leq d\}$, then any polynomial of degree at most d can be written as $f(\mathbf{x}) = \sum_{\boldsymbol{\alpha}\in\Gamma_d^n} b_{\boldsymbol{\alpha}}\mathbf{x}^{\boldsymbol{\alpha}}$. The support of f is defined by $\text{supp}(f) = \{\boldsymbol{\alpha}\in\Gamma_d^n : b_{\boldsymbol{\alpha}} \neq 0\}$. Let $u(\mathbf{x}, \Gamma_d^n)$, be a column vector with the monomials $\mathbf{x}^{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}\in\Gamma_d^n$. The size of the vector $u\big(\mathbf{x}, \Gamma_d^n\big)$ is equal to $\binom{n+d}{d} = \frac{(n+d)!}{n!d!}$, and will be denoted by $g(n, d)$. We will assume without loss of generality that this vector has the following structure

$$\begin{aligned} u(\mathbf{x}, \Gamma_d^n) = \big[&1, x_1, x_2, \ldots, x_n, x_1^2, x_1x_2, \ldots, \\ & x_2^2, x_2x_3, \ldots, x_n^2, \ldots, x_n^{d-1}, x_1^d, x_1^{d-1}x_2, \ldots, x_n^d\big]^\top. \end{aligned}$$

Remark 2.1. Note that $u\big(\mathbf{x}, \Gamma_d^n\big)$ can be written $u\big(\mathbf{x}, \Gamma_d^n\big)^\top = [u\big(\mathbf{x}, \Gamma_{d-1}^n\big)^\top, x_1^d, x_1^{d-1}x_2, \ldots, x_n^d]$.

If $Q\in\mathbb{R}^{r_1\times r_2}$ is a matrix, then the element in position (i, j) will be denoted by $[Q]_{i,j}$ (if $r_1 = 1$ or $r_2 = 1$, the ith element of the vector will be denoted by $[Q]_i$). Likewise, if $Q_1, Q_2 \in \mathbb{R}^{r_1\times r_2}$ are two matrices we will use the usual inner product $\langle Q_1, Q_2\rangle = \sum_{1\leq i\leq r_1}\sum_{1\leq j\leq r_2}[Q_1]_{i,j}[Q_2]_{i,j}$ and its induced norm $\|Q\|^2 = \langle Q, Q\rangle$. $\text{Diag}(x_1, x_2, \ldots, x_n)$ is the function returning a diagonal matrix of dimensions $n\times n$ with x_i in the entry (i, i) for $i = 1,2,\ldots,n$. For any symmetric matrix $Q\in\mathbb{R}^{r\times r}$, $Q\succeq 0$ ($\succ 0$) means that Q is positive semidefinite (resp., definite). For any symmetric matrix $Q\in\mathbb{R}^{r\times r}$ define $\lambda_i(Q)$ as the ith largest eigenvalue of Q, i.e., $\lambda_1(Q) \leq \lambda_2(Q) \leq \cdots \leq \lambda_r(Q)$. For any symmetric matrix $Q\in\mathbb{R}^{r\times r}$, denote $\Omega_Q\in\mathbb{R}^{r\times r}$ as the matrix such that $Q = \Omega_Q \text{Diag}(\lambda_1(Q), \lambda_2(Q), \ldots, \lambda_r(Q))\Omega_Q^\top$ (eigenvalue decomposition). Finally, define $\Lambda(Q, \epsilon)$, as the number of eigenvalues of the symmetric matrix Q that are smaller than $\epsilon\in\mathbb{R}$.

3. SDP relaxations for POP

We use the relaxations formulated in Lasserre (2001) to find an approximate solution for problem (2). This section briefly describes such relaxations for constrained polynomial problems and studies some of their properties.

3.1. Lasserre hierarchy

Consider the POP

$$\begin{aligned} p^\star := \min_{\mathbf{x}\in\mathbb{R}^n} & f(\mathbf{x}) \\ s.t.\ & h_i(\mathbf{x}) \geq 0,\ i = 1,2,\ldots,m, \end{aligned} \tag{2}$$

where f and h_i $(i = 1,2,\ldots,m)$, are n-dimensional polynomial functions with degrees $d, d_1, d_2, \ldots, d_m$, respectively. Writing $f(\mathbf{x}) = \sum_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}\mathbf{x}^{\boldsymbol{\alpha}}$ and since $zz^\top$ is always positive semidefinite for any real vector, we can obtain the following equivalent problem,

$$\begin{aligned} p^\star : = \min_{\mathbf{x}\in\mathbb{R}^n} & \sum_{\boldsymbol{\alpha}\in\mathcal{F}^w} b_{\boldsymbol{\alpha}}\mathbf{x}^{\boldsymbol{\alpha}} \\ s.t.\ & u(\mathbf{x}, \Gamma_w^n)u(\mathbf{x}, \Gamma_w^n)^\top \succeq 0, \\ & u(\mathbf{x}, \Gamma_{w-\tilde{d}_i}^n)u(\mathbf{x}, \Gamma_{w-\tilde{d}_i}^n)^\top h_i(\mathbf{x}) \succeq 0,\ i = 1,2,\ldots,m, \end{aligned} \tag{3}$$

where $\mathcal{F}^w = \Gamma_{2w}^n \setminus \{[0,0,\ldots,0]^\top\}$, $\tilde{d} = \lceil d/2\rceil$, $\tilde{d}_i = \lceil d_i/2\rceil$ $(i = 1,2,\ldots,m)$, and w a positive integer such that $w \geq w_{\min}$ with $w_{\min} = \max\{\tilde{d}, \tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_m\}$. Replacing the monomial $\mathbf{x}^{\boldsymbol{\alpha}}$ by the real variable $y_{\boldsymbol{\alpha}}$ we obtain the Lasserre wth order relaxation

$$\begin{aligned} \sigma_w : = \inf_y & \sum_{\boldsymbol{\alpha}\in\mathcal{F}^w} b_{\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}} \\ s.t.\ & M_w(y) \succeq 0, \\ & M_{w-\tilde{d}_i}(h_i y) \succeq 0,\ i = 1,2,\ldots,m, \end{aligned} \tag{4}$$

where $M_w(y)$ and $M_{w-\tilde{d}_i}(h_i y)$ $(i = 1,2,\ldots,m)$ are the square matrices obtained by replacing all the monomials $\mathbf{x}^{\boldsymbol{\alpha}}$ by the real variable $y_{\boldsymbol{\alpha}}$ in $u(\mathbf{x}, \Gamma_w^n)u(\mathbf{x}, \Gamma_w^n)^\top$ and $u(\mathbf{x}, \Gamma_{w-\tilde{d}_i}^n)u(\mathbf{x}, \Gamma_{w-\tilde{d}_i}^n)^\top h_i(\mathbf{x})$, respectively. The matrices $M_w(y)$ and $M_{w-\tilde{d}_i}(h_i y)$ are called the moment matrix of order w and the localizing matrix, respectively.

The dual of this problem can be written as

$$\begin{aligned} \sigma_w^d : = \sup_{X, Z_i} & -[X]_{1,1} - \sum_{i=1}^m h_i(0)[Z_i]_{1,1} \\ s.t.\ & \langle A_{\boldsymbol{\alpha}}^w, X\rangle + \sum_{i=1}^m \big\langle B_{i,\boldsymbol{\alpha}}^w, Z_i\big\rangle = b_{\boldsymbol{\alpha}},\ \boldsymbol{\alpha}\in\mathcal{F}^w, \\ & X, Z_i \succeq 0,\ i = 1,2,\ldots,m, \end{aligned} \tag{5}$$

where $h_i(0)$ is the monomial of degree zero in the polynomial function h_i, i.e., the constant term, and the matrices $A_{\boldsymbol{\alpha}}^w$ and $B_{i,\boldsymbol{\alpha}}^w$ are such that $M_w(y) = \sum_{\boldsymbol{\alpha}\in\Gamma_{2w}^n} A_{\boldsymbol{\alpha}}^w y_{\boldsymbol{\alpha}}$, and $M_{w-\tilde{d}_i}(h_i y) = \sum_{\boldsymbol{\alpha}\in\Gamma_{2w}^n} B_{i,\boldsymbol{\alpha}}^w y_{\boldsymbol{\alpha}}$, with $y_{\boldsymbol{\alpha}} = 1$ for $\boldsymbol{\alpha} = [0,0,\ldots,0]^\top$.

It is possible to prove that under some assumptions over the feasible set $\{\mathbf{x}\in\mathbb{R}^n : h_i(\mathbf{x}) \geq 0, i = 1,2,\ldots,m\}$, the difference between the optimal value $p^\star$ and σ_w tends to zero as the level of the relaxation w increases. The next theorem formalizes this idea.

Theorem 3.1. *Assume that $K = \{\mathbf{x} \in \mathbb{R}^n : h_i(\mathbf{x}) \ge 0, i = 1, 2, \dots, m\}$ is compact and there exits a real-valued polynomial $v(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R}$ such that $\{\mathbf{x} : v(\mathbf{x}) \ge 0\}$ is compact, and*

$$v(\mathbf{x}) = v_0(\mathbf{x}) + \sum_{i=1}^{m} h_i(\mathbf{x}) v_i(\mathbf{x}) \text{ for all } \mathbf{x} \in \mathbb{R}^n,$$

where the polynomials $v_i(\mathbf{x})$ are all sum of squares, $i = 0, 1, 2, \dots, m$. Then,

(a) Lasserre (2001) *As $w \to \infty$ one has that $\sigma_w \to p^\star$. Moreover, for w sufficiently large, there is no duality gap between problems (4) and (5) if K has a non-empty interior.*
(b) Schweighofer (2005) *If the POP (2) has a unique minimizer $\mathbf{x}^\star = [x_1^\star, x_2^\star, \dots, x_n^\star]^\top$ and $y^w = \{y_{\boldsymbol{\alpha}}^w\}_{\boldsymbol{\alpha} \in \mathcal{F}^w}$ is a solution of the primal SDP relaxation (4), then as $w \to \infty$ one has that $y_{e_j}^w \to x_j^\star$, where $e_j \in \mathbb{R}^n$ is the unit vector with 1 in position j.*

Proof.

(a) See Theorem 4.2 in Lasserre (2001).
(b) See Corollary 3.5 in Schweighofer (2005). □

Remark 3.1. Although the result above guarantees convergence as w tends to infinity, in practice it is very common to get the solution of the POP using a small value of w and in some cases finite convergence can be proved (see for example Lasserre, 2002 for finite convergence in {0, 1} POPs, Lasserre, 2009 and De Klerk & Laurent, 2011 for finite convergence in the convex case, and Nie, 2014 for the general non-linear case).

$$u(\mathbf{x}, \Gamma_{w+1}^n) u(\mathbf{x}, \Gamma_{w+1}^n)^\top = [u(\mathbf{x}, \Gamma_w^n)^\top, x_1^{w+1}, \dots, x_n^{w+1}]^\top [u(\mathbf{x}, \Gamma_w^n)^\top, x_1^{w+1}, \dots, x_n^{w+1}]$$
$$= \begin{bmatrix} u(\mathbf{x}, \Gamma_w^n) u(\mathbf{x}, \Gamma_w^n)^\top & u(\mathbf{x}, \Gamma_w^n)[x_1^{w+1}, \dots, x_n^{w+1}] \\ [x_1^{w+1}, \dots, x_n^{w+1}]^\top u(\mathbf{x}, \Gamma_w^n)^\top & [x_1^{w+1}, \dots, x_n^{w+1}]^\top [x_1^{w+1}, \dots, x_n^{w+1}] \end{bmatrix} \quad (6)$$

3.2. Properties of the SDP relaxations

This section studies the properties of the SDP relaxations (4) and (5). In particular, we want to relate the parameters $A_{\boldsymbol{\alpha}}^{w-1}$ and $B_{i,\boldsymbol{\alpha}}^{w-1}$ $(i = 1, 2, \dots, m)$ for different values of w. To understand the relation between two levels in the hierarchy consider the following example.

Example 3.1. Let $f(\mathbf{x}) = 4x^2 - 2x$ and $h_1(\mathbf{x}) = 3 - x^2$. In this case $d = d_1 = 2$. The moment and localizing moment matrices for $w = 1$ and $w = 2$ are:

- M_w: $M_1(y) = \begin{bmatrix} 1 & y_{[1]} \\ y_{[1]} & y_{[2]} \end{bmatrix}, M_2(y) = \begin{bmatrix} 1 & y_{[1]} & y_{[2]} \\ y_{[1]} & y_{[2]} & y_{[3]} \\ y_{[2]} & y_{[3]} & y_{[4]} \end{bmatrix}$.
- $M_{w-1}(h_1 y)$:
$M_0(h_1 y) = [3 - y_{[2]}], M_1(h_1 y) = \begin{bmatrix} 3 - y_{[2]} & 3y_{[1]} - y_{[3]} \\ 3y_{[1]} - y_{[3]} & 3y_{[2]} - y_{[4]} \end{bmatrix}$.

Then, it is easy to see that the first and second order SDP relaxations are given by the following parameters:

- $b_{\boldsymbol{\alpha}}$: $b_{[1]} = -2$, $b_{[2]} = 4$, $b_{\boldsymbol{\alpha}} = 0$ if $\boldsymbol{\alpha} \notin \{[1], [2]\}$.
- $A_{\boldsymbol{\alpha}}^1$: $A_{[0]}^1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, A_{[1]}^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, A_{[2]}^1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.
- $A_{\boldsymbol{\alpha}}^2$: $A_{[0]}^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A_{[1]}^2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, A_{[2]}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$,
$A_{[3]}^2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, A_{[4]}^2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.
- $B_{\boldsymbol{\alpha}}^1$: $B_{1,[0]}^1 = [3], B_{1,[1]}^1 = [0], B_{1,[2]}^1 = [-1]$.
- $B_{\boldsymbol{\alpha}}^2$: $B_{1,[0]}^2 = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}, B_{1,[1]}^2 = \begin{bmatrix} 0 & 3 \\ 3 & 0 \end{bmatrix}, B_{1,[2]}^2 = \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix}$,
$B_{1,[3]}^2 = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}, B_{1,[4]}^2 = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$.

Note that the matrices $A_{[1]}^1$ and $A_{[2]}^1$ are the 2nd order leading principal sub-matrices of the matrices $A_{[1]}^2$ and $A_{[2]}^2$, respectively. Similarly, $B_{1,[0]}^1$, $B_{1,[1]}^1$ and $B_{1,[2]}^1$ are the 1st order leading principal sub-matrices of the matrices $B_{1,[0]}^2$, $B_{1,[1]}^2$ and $B_{1,[2]}^2$, respectively. Also, the entries of the 2nd and 1st order leading principal sub-matrices of $A_{[3]}^2$, $A_{[4]}^2$, and $B_{1,[3]}^2$, $B_{1,[4]}^2$, respectively, are all zero. Finally, notice that $b_{\boldsymbol{\alpha}} = 0$ for any $\boldsymbol{\alpha}$ such that $\sum_{i=1}^n \alpha_i > 2$. The next lemma formalizes the observations made above.

Lemma 3.1. *If $\tilde{w} \ge w_{\min}$, then the SDP relaxations (4) and (5) of order $w = \tilde{w}$ and $w = \tilde{w} + 1$ satisfy:*

(a) *$b_{\boldsymbol{\alpha}} = 0$ for any $\boldsymbol{\alpha}$ such that $\sum_{i=1}^n \alpha_i > 2w_{\min}$.*
(b) *For any $\boldsymbol{\alpha} \in \Gamma_{2\tilde{w}}^n$ the $g(n, \tilde{w})$th order leading principal sub-matrix of $A_{\boldsymbol{\alpha}}^{\tilde{w}+1}$ is equal to $A_{\boldsymbol{\alpha}}^{\tilde{w}}$, and the $g(n, \tilde{w} - \tilde{d}_i)^{th}$ order leading principal sub-matrix of $B_{i,\boldsymbol{\alpha}}^{\tilde{w}+1}$ is equal to $B_{i,\boldsymbol{\alpha}}^{\tilde{w}}$ for $i = 1, 2, \dots, m$.*
(c) *For any $\boldsymbol{\alpha} \in \Gamma_{2\tilde{w}+1}^n \setminus \Gamma_{2\tilde{w}}^n$, the entries of the $g(n, \tilde{w})$th order leading principal sub-matrix of $A_{\boldsymbol{\alpha}}^{\tilde{w}+1}$ and the $g(n, \tilde{w} - \tilde{d}_i)$th order leading principal sub-matrix of $B_{i,\boldsymbol{\alpha}}^{\tilde{w}+1}$ $(i = 1, 2, \dots, m)$, are equal to zero.*

Proof.

(a) Given that the degree of f is d and $2w_{\min} \ge d$, any monomial of degree greater than $2w_{\min}$ must have a zero coefficient.
To prove (b) and (c), first note that according to Remark 2.1 we have

Also, we constructed $M_w(y)$ and $M_{w-\tilde{d}_i}(h_i y)$ by replacing every monomial $\mathbf{x}^{\boldsymbol{\alpha}}$ for the real variable $y_{\boldsymbol{\alpha}}$ in $u(\mathbf{x}, \Gamma_{w+1}^n) u(\mathbf{x}, \Gamma_{w+1}^n)^\top$ and $u(\mathbf{x}, \Gamma_{w-\tilde{d}_i}^n) u(\mathbf{x}, \Gamma_{w-\tilde{d}_i}^n)^\top h_i(\mathbf{x})$, respectively; and that $A_{\boldsymbol{\alpha}}^w$, $B_{i,\boldsymbol{\alpha}}^w$ are such that $M_w(y) = \sum_{\boldsymbol{\alpha} \in \Gamma_{2w}^n} A_{\boldsymbol{\alpha}}^w y_{\boldsymbol{\alpha}}$ and $M_{w-\tilde{d}_i}(h_i y) = \sum_{\boldsymbol{\alpha} \in \Gamma_{2w}^n} B_{i,\boldsymbol{\alpha}}^w y_{\boldsymbol{\alpha}}$. Using these facts and Eq. (6) we have that,

$$M_{\tilde{w}+1}(y) = \sum_{\boldsymbol{\alpha} \in \Gamma_{2(\tilde{w}+1)}^n} A_{\boldsymbol{\alpha}}^{\tilde{w}+1} y_{\boldsymbol{\alpha}} = \begin{bmatrix} M_{\tilde{w}}(y) & Q_1(y) \\ Q_1(y)^\top & Q_2(y) \end{bmatrix}$$
$$= \begin{bmatrix} \sum_{\boldsymbol{\alpha} \in \Gamma_{2\tilde{w}}^n} A_{\boldsymbol{\alpha}}^{\tilde{w}} y_{\boldsymbol{\alpha}} & Q_1(y) \\ Q_1(y)^\top & Q_2(y) \end{bmatrix}, \quad (7)$$

where $Q_1(y)$ and $Q_2(y)$ are the matrices obtained by replacing the monomials $\mathbf{x}^{\boldsymbol{\alpha}}$ for the real variable $y_{\boldsymbol{\alpha}}$ in the matrices $[x_1^{w+1}, \dots, x_n^{w+1}]^\top u(\mathbf{x}, \Gamma_w^n)^\top$ and $[x_1^{w+1}, \dots, x_n^{w+1}]^\top [x_1^{w+1}, \dots, x_n^{w+1}]$, respectively. Using the same reasoning, we obtain

$$M_{(w+1)-\tilde{d}_i}(h_i y) = \sum_{\boldsymbol{\alpha} \in \Gamma_{2(\tilde{w}+1)}^n} B_{i,\boldsymbol{\alpha}}^{\tilde{w}+1} y_{\boldsymbol{\alpha}} = \begin{bmatrix} \sum_{\boldsymbol{\alpha} \in \Gamma_{2\tilde{w}}^n} B_{i,\boldsymbol{\alpha}}^{\tilde{w}} y_{\boldsymbol{\alpha}} & \tilde{Q}_1(y) \\ \tilde{Q}_1(y)^\top & \tilde{Q}_2(y) \end{bmatrix}, \quad (8)$$

where again the matrices $\tilde{Q}_1(y)$ and $\tilde{Q}_2(y)$ are obtained by replacing the monomials $\mathbf{x}^{\boldsymbol{\alpha}}$ by the real variable $y_{\boldsymbol{\alpha}}$ in the matrix $[x_1^{(\tilde{w}+1)-\tilde{d}_i}, \dots, x_n^{(\tilde{w}+1)-\tilde{d}_i}]^\top u(\mathbf{x}, \Gamma_{\tilde{w}-\tilde{d}_i}^n)^\top h_i(\mathbf{x})$ and the matrix $[x_1^{(\tilde{w}+1)-\tilde{d}_i}, \dots, x_n^{(\tilde{w}+1)-\tilde{d}_i}]^\top [x_1^{(\tilde{w}+1)-\tilde{d}_i}, \dots, x_n^{(\tilde{w}+1)-\tilde{d}_i}] h_i(\mathbf{x})$, respectively.

(b) Using Eqs. (7), (8), is easy to see that $\sum_{\boldsymbol{\alpha}\in\Gamma^n_{2\tilde{w}}} A^{\tilde{w}}_{\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}}$ and $\sum_{\boldsymbol{\alpha}\in\Gamma^n_{2\tilde{w}}} B^{\tilde{w}}_{i,\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}}$ correspond to the $g(n,\tilde{w})$th and $g(n,\tilde{w}-\tilde{d}_i)$th order leading principal sub-matrices of $\sum_{\boldsymbol{\alpha}\in\Gamma^n_{2(\tilde{w}+1)}} A^{\tilde{w}+1}_{\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}}$ and $\sum_{\boldsymbol{\alpha}\in\Gamma^n_{2(\tilde{w}+1)}} B^{\tilde{w}+1}_{i,\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}}$, respectively, from where statement (b) follows.

(c) Notice that $\sum_{\boldsymbol{\alpha}\in\Gamma^n_{2\tilde{w}}} A^{\tilde{w}}_{\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}}$ does not contain any $y_{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}: \sum \alpha_j > 2\tilde{w}+1$ (or equivalently, any $y_{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}\in\Gamma^n_{2\tilde{w}+1}\setminus\Gamma^n_{2\tilde{w}}$ is multiplied by a zero matrix). Given statement (b), we can conclude then that the $g(n,\tilde{w})$th order leading principal sub-matrix of $A^{\tilde{w}+1}_{\boldsymbol{\alpha}}$ is zero for any $\boldsymbol{\alpha}\in\Gamma^n_{2\tilde{w}+1}\setminus\Gamma^n_{2\tilde{w}}$. A similar argument can be made for the $g(n,\tilde{w}-\tilde{d}_i)$th order leading principal sub-matrix of $B^{\tilde{w}+1}_{i,\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}\in\Gamma^n_{2\tilde{w}+1}\setminus\Gamma^n_{2\tilde{w}}$. □

4. Prolongation operators

Given $\tilde{w}\geq w_{\min} = \max\{\tilde{d},\tilde{d}_1,\tilde{d}_2,\dots,\tilde{d}_m\}$, we will define prolongation operators to relate any point in the SDP relaxations (4) and (5) of order $w=\tilde{w}$, to the SDP relaxation of order $w=\tilde{w}+1$. We will refer to the $\tilde{w}$th SDP space relaxation problem and variables as the coarse problem (or coarse relaxation) and coarse variables. Similarly, we will refer to the $(\tilde{w}+1)$th SDP space relaxation problem and variables as the fine problem (or fine relaxation) and fine variables.

For any order $w\geq w_{\min}$, we will denote the primal variables for the wth order relaxation (4) as $y^w\in\mathbb{R}^{|\mathcal{F}^w|}$ with $y^w=\{y^w_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha}\in\mathcal{F}^w}$, and the dual variables of the relaxation (5) as $X^w\in\mathbb{R}^{g(n,w)\times g(n,w)}$, and $Z^w\in\mathbb{R}^{g(n,w-\tilde{d}_1)\times g(n,w-\tilde{d}_1)}\times\dots\times\mathbb{R}^{g(n,w-\tilde{d}_m)\times g(n,w-\tilde{d}_m)}$ with $Z^w=\left(Z^w_1,Z^w_2,\dots,Z^w_m\right)$.

For (y^w,X^w,Z^w) we define the dual residuals at the point (X^w,Z^w) $(r^w_{\boldsymbol{\alpha}}(X^w,Z^w))$ as

$$r^w_{\boldsymbol{\alpha}}(X^w,Z^w) := \langle A^w_{\boldsymbol{\alpha}},X^w\rangle + \sum_{i=1}^m \left\langle B^w_{i,\boldsymbol{\alpha}},Z^w_i\right\rangle - b_{\boldsymbol{\alpha}}, \tag{9}$$

for $\boldsymbol{\alpha}\in\mathcal{F}^w$.

4.1. Primal prolongation operator

By inspecting the hierarchy, we notice that the number of primal and dual matrices do not change from the coarse to the fine relaxations. Instead the matrix dimensions increase from one level to the next. For the primal variables we will define a non-linear operator.

Let $\mathrm{proj}_K(\mathbf{x})$ be the projection operator onto the set $K=\{\mathbf{x}\in\mathbb{R}^n : h_i(\mathbf{x})\geq 0,\ i=1,2,\dots,m\}$, i.e.,

$$\mathrm{proj}_K(\mathbf{x}) := \arg\min_{\mathbf{z}\in K}\|\mathbf{x}-\mathbf{z}\|^2, \tag{10}$$

and define $\Pi^w:\mathbb{R}^n\mapsto\mathbb{R}^{|\mathcal{F}^{w+1}|}$ as

$$[\Pi^w(\mathbf{x})]_{\boldsymbol{\alpha}} := x_1^{\alpha_1}x_2^{\alpha_2}\dots x_n^{\alpha_n},\ \boldsymbol{\alpha}\in\mathcal{F}^{w+1}, \tag{11}$$

for any $\mathbf{x}\in\mathbb{R}^n$.

Using Eqs. (10) and (11), we define a non-linear operator $P^w_y:\mathbb{R}^{|\mathcal{F}^w|}\mapsto\mathbb{R}^{|\mathcal{F}^{w+1}|}$ for any primal point $y^w\in\mathbb{R}^{|\mathcal{F}^w|}$ by

$$P^w_y(y^w) := \Pi^w\left(\mathrm{proj}_K\left(\left[y^w_{e_1},y^w_{e_2},\dots,y^w_{e_n}\right]^\top\right)\right) \tag{12}$$

where $e_j\in\mathbb{R}^n$ is a unit vector with 1 in position j.

Theorem 4.1. *Let $\tilde{w}\geq w_{\min}$ and $y^{\tilde{w}}$ a point (not necessarily feasible) of the SDP relaxation of order $w=\tilde{w}$ defined in (4). If $y^{\tilde{w}+1}=P^{\tilde{w}}_y(y^{\tilde{w}})$ is defined according to prolongation operator (12) for $w=\tilde{w}$, then $y^{\tilde{w}+1}$ is feasible for the primal SDP relaxation (4) of order $w=\tilde{w}+1$.*

Proof. To prove that $M_{\tilde{w}+1}(P^{\tilde{w}}_y(y^{\tilde{w}}))$ is positive semidefinite, notice that for any $\mathbf{x}\in\mathbb{R}^n$ we have that $M_{\tilde{w}+1}(\Pi^{\tilde{w}}(\mathbf{x})) = u([\mathbf{x},\Gamma^n_{2(\tilde{w}+1)}])u([\mathbf{x},\Gamma^n_{2(\tilde{w}+1)}])^\top$, we can conclude that $M_{\tilde{w}+1}(P^{\tilde{w}}_y(y^{\tilde{w}})) = M_{\tilde{w}+1}(\Pi^{\tilde{w}}(\mathrm{proj}_K([y^w_{e_1},y^w_{e_2},\dots,y^w_{e_n}]^\top)))\succeq 0$ by using the fact that $\mathbf{z}\mathbf{z}^\top$ is positive semidefinite for any real vector $\mathbf{z}$.

Similarly, to prove the positive semidefiniteness of the localizing matrices notice that for any $\mathbf{x}\in\mathbb{R}^n$ we have $M_{(\tilde{w}+1)-\tilde{d}_i}(h_i\Pi^{\tilde{w}}(\mathbf{x})) = M_{(\tilde{w}+1)-\tilde{d}_i}(\Pi^{\tilde{w}}(\mathbf{x}))h_i(\mathbf{x})$. Therefore, given that $M_{(\tilde{w}+1)-\tilde{d}_i}(P^{\tilde{w}}_y(y^{\tilde{w}}))$ is positive semidefinite (we can write it as $u([\mathbf{x},\Gamma^n_{2(\tilde{w}+1)}])u([\mathbf{x},\Gamma^n_{2(\tilde{w}+1)}])^\top$ with $\mathbf{x}=\mathrm{proj}_K([y^w_{e_1},y^w_{e_2},\dots,y^w_{e_n}]^\top)$), and $h_i(\mathrm{proj}_K([y^w_{e_1},y^w_{e_2},\dots,y^w_{e_n}]^\top))\geq 0$ (the projection over K guarantees this), we can conclude that $M_{(\tilde{w}+1)-\tilde{d}_i}(h_i\Pi^{\tilde{w}}(\mathrm{proj}_K([y^w_{e_1},y^w_{e_2},\dots,y^w_{e_n}]^\top))$ is positive semidefinite. □

4.2. Dual prolongation operator

As already mentioned, the number of dual matrices in the coarse and fine relaxations is m, i.e., the number constraints in the dual relaxation, but the size of the matrices is larger in the fine problem. In this case, the prolongation will be constructed by using the coarse matrices as the leading principal sub-matrices of the fine matrices. In particular, for any $w\geq w_{\min} = \max\{\tilde{d},\tilde{d}_1,\tilde{d}_2,\dots,\tilde{d}_m\}$ let $P^w_X:\mathbb{R}^{g(n,w)\times g(n,w)}\mapsto\mathbb{R}^{g(n,w+1)\times g(n,w+1)}$ be the prolongation operator for the coarse variable X^w, and $P^w_Z:\mathbb{R}^{g(n,w-\tilde{d}_1)\times g(n,w-\tilde{d}_1)}\times\dots\times\mathbb{R}^{g(n,w-\tilde{d}_m)\times g(n,w-\tilde{d}_m)}\mapsto\mathbb{R}^{g(n,(w+1)-\tilde{d}_1)\times g(n,(w+1)-\tilde{d}_1)}\times\dots\times\mathbb{R}^{g(n,(w+1)-\tilde{d}_m)\times g(n,(w+1)-\tilde{d}_m)}$ be the prolongation operator for the coarse variable Z^w. If $X^{w+1}=P^w_X(X^w)$ and $Z^{w+1}=(Z^{w+1}_1,Z^{w+1}_2,\dots,Z^{w+1}_m)=P^w_Z(Z^w)$ then

$$X^{w+1}=P^w_X(X^w)=\begin{bmatrix}X^w & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}, \tag{13}$$

$$Z^{w+1}_i=[P^w_Z(Z^w)]_i=\begin{bmatrix}Z^w_i & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix},\ i=1,2,\dots,m, \tag{14}$$

where $\mathbf{0}$'s are zero matrices of appropriate size. The next theorems characterize the feasibility of any prolongated coarse dual point (X^w,Z^w).

Theorem 4.2. *Let $\tilde{w}\geq w_{\min}$ and $(X^{\tilde{w}},Z^{\tilde{w}})$ a point (not necessarily feasible) of the dual SDP relaxation (5) of order $w=\tilde{w}$. If $X^{\tilde{w}+1}=P^{\tilde{w}}_X(X^{\tilde{w}})$ and $Z^{\tilde{w}+1}=P^{\tilde{w}}_Z(Z^{\tilde{w}})$ are defined according to Eqs. (13) and (14) with $w=\tilde{w}$ respectively, then for any $\boldsymbol{\alpha}\in\mathcal{F}^{\tilde{w}+1}$ we have*

$$r^{\tilde{w}+1}_{\boldsymbol{\alpha}}(X^{\tilde{w}+1},Z^{\tilde{w}+1}) = \begin{cases} r^{\tilde{w}}_{\boldsymbol{\alpha}}(X^{\tilde{w}},Z^{\tilde{w}}), & \text{if } \boldsymbol{\alpha}\in\mathcal{F}^{\tilde{w}},\\ 0, & \textit{otherwise,}\end{cases} \tag{15}$$

where $r^{\tilde{w}}_{\boldsymbol{\alpha}}(X^{\tilde{w}},Z^{\tilde{w}})$ is the dual residual defined in Eq. (9).

Proof. Note that

$$\begin{aligned}\left\langle A^{\tilde{w}+1}_{\boldsymbol{\alpha}},X^{\tilde{w}+1}\right\rangle &= \sum_{1\leq i,j\leq g(n,\tilde{w}+1)} [A^{\tilde{w}+1}_{\boldsymbol{\alpha}}]_{i,j}[X^{\tilde{w}+1}]_{i,j}\\ &= \sum_{1\leq i,j\leq g(n,\tilde{w}+1)} [A^{\tilde{w}+1}_{\boldsymbol{\alpha}}]_{i,j}[P^{\tilde{w}}_X(X^{\tilde{w}})]_{i,j}\\ &= \sum_{1\leq i,j\leq g(n,\tilde{w})} [A^{\tilde{w}+1}_{\boldsymbol{\alpha}}]_{i,j}[X^{\tilde{w}}]_{i,j}\\ &= \sum_{1\leq i,j\leq g(n,\tilde{w})} [A^{\tilde{w}}_{\boldsymbol{\alpha}}]_{i,j}[X^{\tilde{w}}]_{i,j} = \left\langle A^{\tilde{w}}_{\boldsymbol{\alpha}},X^{\tilde{w}}\right\rangle,\end{aligned}$$

where we used the fact that according to Eq. (13), $[X^{\tilde{w}+1}_k]_{i,j} = [P^{\tilde{w}}_X(X^{\tilde{w}})]_{i,j}=0$ for any $i,j>g(n,\tilde{w})$, and Lemma 3.1 (b) to replace $A^{\tilde{w}+1}_{\boldsymbol{\alpha}}$ by $A^{\tilde{w}}_{\boldsymbol{\alpha}}$. Similarly, using Eq. (14) and the second part of Lemma 3.1 (b) we can deduce that $\left\langle B^{\tilde{w}+1}_{i,\boldsymbol{\alpha}},Z^{\tilde{w}+1}_i\right\rangle = \left\langle B^{\tilde{w}}_{i,\boldsymbol{\alpha}},Z^{\tilde{w}}_i\right\rangle$. Then, if $\boldsymbol{\alpha} \in \mathcal{F}^{\tilde{w}}$, we can write $r_{\boldsymbol{\alpha}}^{\tilde{w}+1}$ as

$$r_{\boldsymbol{\alpha}}^{\tilde{w}+1}(X^{\tilde{w}+1}, Z^{\tilde{w}+1}) = \left\langle A_{\boldsymbol{\alpha}}^{\tilde{w}+1}, X^{\tilde{w}+1}\right\rangle + \sum_{i=1}^{m}\left\langle B_{i,\boldsymbol{\alpha}}^{\tilde{w}+1}, Z_i^{\tilde{w}+1}\right\rangle - b_{\boldsymbol{\alpha}}$$
$$= \left\langle A_{\boldsymbol{\alpha}}^{\tilde{w}}, X^{\tilde{w}}\right\rangle + \sum_{i=1}^{m}\left\langle B_{i,\boldsymbol{\alpha}}^{\tilde{w}}, Z_i^{\tilde{w}}\right\rangle - b_{\boldsymbol{\alpha}} = r_{\boldsymbol{\alpha}}^{\tilde{w}}(X^{\tilde{w}}, Z^{\tilde{w}}).$$

Likewise, if $\boldsymbol{\alpha} \notin \mathcal{F}^{\tilde{w}}$, then $b_{\boldsymbol{\alpha}} = 0$ as $\sum_i \alpha_i > 2w_{\min}$ (Lemma 3.1 (a)), and $\left\langle A_{\boldsymbol{\alpha}}^{\tilde{w}+1}, X^{\tilde{w}+1}\right\rangle = 0$ and $\left\langle B_{i,\boldsymbol{\alpha}}^{\tilde{w}+1}, Z_i^{\tilde{w}+1}\right\rangle = 0$ for any $i = 1, 2, \ldots, m$ (Lemma 3.1 (c)). Hence,

$$r_{\boldsymbol{\alpha}}^{\tilde{w}+1}(X^{\tilde{w}+1}, Z^{\tilde{w}+1})$$
$$= \left\langle A_{\boldsymbol{\alpha}}^{\tilde{w}+1}, X^{\tilde{w}+1}\right\rangle + \sum_{i=1}^{m}\left\langle A_{i,\boldsymbol{\alpha}}^{\tilde{w}+1}, Z_i^{\tilde{w}+1}\right\rangle - b_{\boldsymbol{\alpha}} = 0 - b_{\boldsymbol{\alpha}} = 0.$$

□

Lemma 4.1. *Under the assumptions of Theorem 4.2, if $(X^{\tilde{w}}, Z^{\tilde{w}})$ is also a feasible point of the dual SDP relaxation of order $w = \tilde{w}$ defined in (5), then*

(a) $X^{\tilde{w}+1}, Z_i^{\tilde{w}+1} \succeq 0$, *for $i = 1, 2, \ldots, m$.*
(b) $r_{\boldsymbol{\alpha}}^{\tilde{w}+1}(X^{\tilde{w}+1}, Z^{\tilde{w}+1}) = 0$ *for any $\boldsymbol{\alpha} \in \mathcal{F}^{\tilde{w}+1}$.*

Proof.

(a) Using the fact that X^w is feasible, we have that $X^{\tilde{w}} \succeq 0$ and therefore if $\mathbf{z} \in \mathbb{R}^{g(n,\tilde{w}+1)}$ we have that

$$\mathbf{z}^\top X^{\tilde{w}+1}\mathbf{z} = \mathbf{z}^\top \begin{bmatrix} X^{\tilde{w}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{z}$$
$$= [z_1, z_2, \ldots, z_{g(n,\tilde{w})}]X^{\tilde{w}}[z_1, z_2, \ldots, z_{g(n,\tilde{w})}]^\top$$
$$\geq 0.$$

Hence, $X_l^{\tilde{w}+1}$ is positive semidefinite. The same argument applies to $Z_i^{\tilde{w}+1}$ for $i = 1, 2, \ldots, m$.

(b) This statement follows by using Theorem 4.2 and noticing that $r_{\boldsymbol{\alpha}}^{\tilde{w}}(X^{\tilde{w}}, Z^{\tilde{w}}) = 0$ for any $\boldsymbol{\alpha} \in \mathcal{F}^{\tilde{w}}$ because $(X^{\tilde{w}}, Z^{\tilde{w}})$ is feasible for the coarse problem. □

4.3. Duality gap of prolongated variables

This section assumes the conditions of Theorem 3.1 are satisfied by the POP in (2). The next result guarantees that the duality gap of the prolongated coarse solutions tends to zero as the order of the relaxation goes to infinity.

Theorem 4.3. *Assume that the POP (2) has a compact feasible set and a unique solution $\mathbf{x}^\star$ with global minimum $p^\star = \sum_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}}(\mathbf{x}^\star)^{\boldsymbol{\alpha}}$. Furthermore, let $w_0 \in \mathbb{N}$ be such that for any $w \geq w_0$ the wth order SDP relaxations defined in problems (4) and (5), are solvable and have zero duality gap (note that w_0 exists according to Theorem 3.1). For $w \geq w_0$, let y^w and (X^w, Z^w) be a primal and a dual optimal solution for the SDP relaxations of order w respectively. If the operators defined in Eqs. (12)–(14) are used to prolongate these solutions to the level $w+1$, then the duality gap of the prolongated points tends to zero as w tends to infinity, i.e.,*

$$\sum_{\boldsymbol{\alpha}\in\mathcal{F}^{w+1}} b_{\boldsymbol{\alpha}}[P_y^w(y^w)]_{\boldsymbol{\alpha}} - \left(-[P_X^w(X^w)]_{1,1} - \sum_{i=1}^{m} h_i(0)[P_Z^w(Z^w)_i]_{1,1}\right)$$
$$\to 0 \text{ as } w \to \infty.$$

Proof. Using the prolongation operators defined in Eqs. (13) and (14), the objective function of the dual relaxation can be written as

$$-[P_X^w(X^w)]_{1,1} - \sum_{i=1}^{m} h_i(0)[P_Z^w(Z^w)_i]_{1,1}$$
$$= -[X^w]_{1,1} - \sum_{i=1}^{m} h_i(0)[Z_i^w]_{1,1}. \tag{16}$$

Hence, using the fact that $\sum_{\boldsymbol{\alpha}\in\mathcal{F}^w} b_{\boldsymbol{\alpha}}^w y^w \to p^\star$ as $w \to \infty$ (Theorem 3.1 (a)), the zero duality gap of the relaxation, and Eq. (16), we can deduce that

$$-[P_X^w(X^w)]_{1,1} - \sum_{i=1}^{m} h_i(0)[P_Z^w(Z^w)_i]_{1,1} \to p^\star \text{ as } w \to \infty. \tag{17}$$

Now, notice that $\text{proj}_K([y_{e_1}^w, \ldots, y_{e_n}^w]^\top) \to \mathbf{x}^\star$ as $w \to \infty$ because $\text{proj}_K(\mathbf{x}^\star) = \mathbf{x}^\star$ and $y_{e_i}^w \to x_i^\star$ as $w \to \infty$ for $i = 1, 2, \ldots, n$ (Theorem 3.1 (b)). Therefore, using Theorem 3.1 (a), Lemma 3.1 (a) and Eq. (11), we have that

$$\sum_{\boldsymbol{\alpha}\in\mathcal{F}^{w+1}} b_{\boldsymbol{\alpha}}[P_y^w(y^w)]_{\boldsymbol{\alpha}} = \sum_{\boldsymbol{\alpha}\in\mathcal{F}^{w}} b_{\boldsymbol{\alpha}}[\Pi^w(\text{proj}_K([y_{e_1}^w, \ldots, y_{e_n}^w]^\top)]_{\boldsymbol{\alpha}}$$
$$\to \sum_{\boldsymbol{\alpha}\in\mathcal{F}^{w}} b_{\boldsymbol{\alpha}}(\mathbf{x}^\star)^{\boldsymbol{\alpha}} = p^\star, \text{ as } w \to \infty. \tag{18}$$

Finally, using Eqs. (17) and (18) we notice that both the primal and dual objective functions evaluated on the prolongated points converges to $p^\star$ as $w \to \infty$ and therefore their difference convergences to zero as $w \to \infty$. □

5. Numerical experiments

Section 4 results suggest that to solve the $(w+1)$th relaxation we can use the operators (12)–(14), along with the solution of the wth relaxation to provide an initial starting point. Like in the previous sections, we will call the relaxation of order w the coarse relaxation or problem and its variables coarse variables. Similarly, the $(w+1)$th SDP relaxation will be referred to as the fine relaxation or problem, with fine variables. According to Theorem 4.1 and Lemma 4.1, the prolongated points have zero infeasibility in the fine level. Theorem 4.3 indicates that, for any $\epsilon > 0$, we can find a w such that the duality gap of the prolongated points is smaller than ϵ. This section illustrates how the operators can be used with an interior point method to solve the $(w+1)$th SDP relaxation.

As indicated in the introduction, our operators assume that the feasible set of the POP is such that calculating the projection of any point onto the set is "easy". Here we consider numerical examples where the only constraints are $\mathbf{x} \in \{0, 1\}^n$ or $\mathbf{x} \in \{-1, 1\}^n$. These constraints can easily be written as polynomials and the projection of any point can be calculated in closed form. For example, the constraint $\mathbf{x} \in \{0, 1\}^n$ is equivalent to $x_i^2 - x_i = 0, i = 1, 2, \ldots n$ (note that these equalities can be replaced by double inequalities), and the projection from box bounds onto the feasible set can be calculated as,

$$\left[\text{proj}_{\{0,1\}^n}(\mathbf{x})\right]_i = \begin{cases} 0, & \text{if } x_i \leq 0.5, \\ 1, & \text{if } x_i > 0.5. \end{cases}$$

When the POP only has integer constraints $\{0, 1\}$ (or $\{-1, 1\}$), the SDP relaxations can be transformed into an equivalent smaller SDP problem (Lasserre, 2002). For $\{0, 1\}$ POPs, the primal SDP relaxation (4) can be reduced by first eliminating the constraints $M_{w-\tilde{d}_i}(h_i y) \succeq 0$ $(i = 1, 2, \ldots, m)$, then replacing every variable $y_{\boldsymbol{\alpha}}$ by the variable $y_{\boldsymbol{\beta}}$ with $\beta_i = 1$ if $\alpha_i \geq 1$, and finally deleting the kth column and row of the resulting moment matrix $M_w(y)$ if $[M_w(y)]_{1,k} = [M_w(y)]_{1,l}$ for some $l < k$ (a similar reduction can be done for the $\{-1, 1\}$ case). Let $\tilde{b}_{\boldsymbol{\alpha}}$ and $\tilde{M}_w(y) \in \gamma_w \times \gamma_w$ be the vector and matrix obtained using the procedure described above. Then the reduced relaxation is given by

$$\sigma_w := \inf_y \sum_{\boldsymbol{\alpha} \in \mathcal{F}^w} \tilde{b}_{\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}} \quad s.t.\ \tilde{M}_w(y) \succeq 0, \tag{19}$$

with the dual

$$\sigma_w^d := \sup_X -[X]_{1,1} \quad s.t.\ \left\langle \tilde{A}_{\boldsymbol{\alpha}}^w, X \right\rangle = b_{\boldsymbol{\alpha}},\ \boldsymbol{\alpha} \in \mathcal{F}^w,\quad X \succeq 0, \tag{20}$$

where $\tilde{M}_w(y) = \sum_{\boldsymbol{\alpha} \in \Gamma_{2w}^n} \tilde{A}_{\boldsymbol{\alpha}}^w y_{\boldsymbol{\alpha}}$.

The result obtained in Lemma 3.1 literal (a) is still valid for these reduced SDP relaxations, and a similar property to Lemma 3.1 literals (b) and (c) can also be proved for the matrices $\tilde{A}_{\boldsymbol{\alpha}}^w$. In particular, if $\tilde{A}_{\boldsymbol{\alpha}}^w$ has dimensions $\gamma_w \times \gamma_w$, then for any $\boldsymbol{\alpha} \in \Gamma_{2\tilde{w}}^n$ we have that $\tilde{A}_{\boldsymbol{\alpha}}^w$ is the γ_w^{th} order leading principal sub-matrix of $\tilde{A}_{\boldsymbol{\alpha}}^{w+1}$; and for any $\boldsymbol{\alpha} \in \Gamma_{2(\tilde{w}+1)}^n \setminus \Gamma_{2\tilde{w}}^n$ the entries of the γ_w^{th} order leading sub-matrix of $\tilde{A}_{\boldsymbol{\alpha}}^{w+1}$ are equal to zero.

The prolongation operators can still be used for these new problems. For the primal relaxation, we use the prolongation defined in Eq. (12). The dual variable X^w will be prolongated using the same idea used in Eq. (13), i.e., if the variable X^w has dimensions $\gamma_w \times \gamma_w$ then $\tilde{P}_X : \mathbb{R}^{\gamma_w \times \gamma_w} \mapsto \mathbb{R}^{\gamma_{w+1} \times \gamma_{w+1}}$ is defined by

$$\tilde{P}_X(X^w) := \begin{bmatrix} X^w & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{21}$$

Notice that all the results proved in Theorems 4.1 and 4.3 and Lemma 4.1, are still valid for the new hierarchy, and therefore, after prolongating a feasible coarse point, the new point is feasible in the fine SDP space, and the duality gap obtained for prolongated coarse optimal points tends to zero as the order of the relaxation gets larger.

As mentioned in Remark 3.1, it is possible to prove that the Lasserre hierarchy has finite convergence for the $\{0, 1\}$ and $\{-1, 1\}$ POP (see Theorem 3.2 in Lasserre, 2002). Furthermore, for all the problems we find the underlying POP solution using relaxation orders $w \geq w_{\min}$ smaller than the ones predicted by the theory. Therefore, to solve the original POP we can solve in a sequential manner the sparse SDP relaxations starting with $w = w_{min} = \max\{\tilde{d}, \tilde{d}_1, \ldots, \tilde{d}_m\}$ and increasing the relaxation order until a solution or approximate solution is found. If a solution is found by solving the SDP relaxation of order $w > w_{\min}$, this procedure implies solving the relaxation of order $w_{\min}, w_{\min}+1, \ldots, w$. The idea is to exploit the information calculated when solving the lower order relaxations to solve the relaxation of order w using the operators defined in the previous section.

Consider the following benchmark test problems:

- Quadratic optimization $\{0, 1\}$: given $l_i \in \mathbb{R}$ $(i = 1, 2, \ldots, n)$ and $k_{i,j} \in \mathbb{R}$ $(1 \leq i, j \leq n)$ the problem is

$$\min_{\mathbf{x} \in \mathbb{R}^n} \sum_{i=1}^n l_i x_i^2 + \sum_{i<j} k_{i,j} x_i x_j \quad s.t.\ x_i^2 - x_i = 0, i = 1, 2, \ldots, n.$$

- MAX-CUT: given a graph $G(V, E)$ with nodes $V = \{V_1, V_2, \ldots, V_n\}$, a set of edges $E = \{(i, j) : 1 \leq i, j, \leq n, \text{if } i \text{ is connected to } j\}$, and a symmetric matrix W with $[W]_{i,j} \neq 0$ if $(i, j) \in E$ and zero otherwise, this problem can be written as

$$\max_{\mathbf{x} \in \mathbb{R}^n} \mathbf{x}^\top L \mathbf{x} \quad s.t.\ x_i^2 = 1, i = 1, 2, \ldots, n,$$

where $L = \mathrm{Diag}([W\mathbf{1}_n]_{1,1}, [W\mathbf{1}_n]_{2,2}, \ldots, [W\mathbf{1}_n]_{n,n}) - W$, and $\mathbf{1}_n \in \mathbb{R}^n$ is a vector of ones.

- Partitioning an integer sequence: given an integer vector $a \in \mathbb{N}^n$, the problem consists in determining if there exists a vector $\mathbf{x} \in \{-1, 1\}^n$ such that $\mathbf{a}^\top \mathbf{x} = 0$, i.e.,

$$\min_{\mathbf{x} \in \mathbb{R}^n} (\mathbf{a}^\top \mathbf{x})^2 \quad s.t.\ x_i^2 = 1, i = 1, 2, \ldots, n.$$

We generate 100 quadratic $\{0, 1\}$ POPs, by selecting the coefficients l_i, $k_{i,j}$ uniformly from the interval $[-1, 1]$ using $n = 10$ and $n = 20$, i.e., a total of 200 problems. Similarly, we generate 100 random MAX-CUT problems selecting the weights $w_{i,j}$ uniformly from the interval [0,1] and another 100 with weights between $[-1, 1]$, using $n = 10$ and $n = 20$, i.e., a total of 400 problems.[1] For the integer partitioning POP we generate 100 sequences of the form $\mathbf{a} = [a_1, a_2, \ldots a_{n/2}, a_1, a_2, \ldots, a_{n/2}]$, by uniformly selecting each $a_1, a_2, \ldots, a_{n/2}$, from the integer set $\{1, 2, \ldots, 100\}$ for $n = 10$ and $n = 14$ (note that the structure of the vector $\mathbf{a}$ guarantees that the problem always has a solution).

We use the MATLAB code SparsePOP version 3.00 (Waki, Kim, Kojima, Muramatsu, & Sugimoto, 2008) to generate the SDP relaxations as well as the MAX-CUT problems (we change the lines 22 and 33 in the file genMAXCUT.m to obtain weights between [0,1] and $[-1, 1]$ as the original code generates integer weights between $[-100, 100]$). The POPs used in this work do not have a unique solution and therefore the results of Theorem 4.3 do not apply. Following Waki et al. (2006), we perturbed the polynomial objective function by adding a small linear term to guarantee a unique solution (see Section 5.1 Waki et al., 2006 for more details), in particular we set the parameter param.perturbation in SparsePOP equal to 10^{-4} for the integer partitioning problem and 10^{-6} for the Quadratic and the MAX-CUT problems. To solve the resulting SDP relaxations we use the infeasible interior point method implemented in SDPT3 version 4.0 (Toh, Todd, & Tütüncü, 2012), SeDuMi version 1.0 (Sturm, 1999), and Mosek version 8.1.0 (we used CVX 2.1 as interface to call Mosek). The tolerance for the three solvers was set to 10^{-7}. All the experiments are done in MATLAB version 2017a in an Intel Core i7-6700 CPU @ 3.40 Gigahertz Ubuntu 16.04 workstation with 16 gigabytes of RAM.

Let y^w be the primal relaxation solution of order w, $\mathbf{y_1}^w = [y_{e_1}^w, \ldots, y_{e_n}^w]$ ($e_j \in \mathbb{R}^n$ is a unit vector with 1 in position j), and $(\mathbf{y_1}^w)^{\boldsymbol{\alpha}} = (y_{e_1}^w)^{\alpha_1}, \ldots, (y_{e_n}^w)^{\alpha_n}$. Then, for each problem we solve the relaxation of order $w = 1, 2, \ldots,$ until we find a relaxation that solves the original POP, i.e., until $\sigma_w = \sum_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}} (\mathbf{y_1}^w)^{\boldsymbol{\alpha}}$ and $\mathbf{y_1}^w$ is feasible for the POP. For the Quadratic and the MAX-CUT problems, we consider a POP solved by the SDP relaxation of order w if $|\sigma_w - \sum_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}} (\mathbf{y_1}^w)^{\boldsymbol{\alpha}}| / \max\{1, |\sum_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}} (\mathbf{y_1}^w)^{\boldsymbol{\alpha}}|\} < 10^{-5}$, and if $\max_i\{|(y_{e_i}^w)^2 - y_{e_i}^w|\} < 10^{-2}$ for the Quadratic POP or $\max_i\{|(y_{e_i}^w)^2 - 1|\} < 10^{-2}$ for the MAX-CUT. For the integer partitioning problem, we scale the problem by using $\mathbf{a}/\|\mathbf{a}\|$ instead of $\mathbf{a}$, and considered a POP solved when $|\sum_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}} (\mathbf{y_1}^w)^{\boldsymbol{\alpha}} - 0| < 10^{-5}$ (here we use the fact that the minimum of the POP is zero) and $\max_i\{|(y_{e_i}^w)^2 - 1|\} < 10^{-2}$.

Table 1 shows that the Quadratic and the MAX-CUT problems required at most the second order relaxation to find a solution of the original POP. The integer partitioning problem with 10 polynomial variables needed the third order relaxation for 36 of the POPs, while for 72 of the problems with 14 variables we calculated at least the fourth order relaxation to find a solution (unfortunately for these relaxations the time needed to solve the SDP problem was larger than 2.5 hours for SeDuMi and 18 hours for SDPT3 making the calculation for all the POPs time consuming). Additionally,

[1] We also used MAX-CUT problems with integer weights for the experiments but did not find any significant change in the results compared with the non-integer case.

Table 1
Number of POPs solved by level of the SDP relaxation (if $(y^w, X^w) \in (\mathbb{R}^{m_w}, \mathbb{R}^{\gamma_w \times \gamma_w})$ then the dimensions of the SDP relaxation m_w, γ_w are in parenthesis), POPs not solved by relaxation of order w but solved by $\text{proj}_K([y^w_{e_1}, \dots, y^w_{e_n}])$.

Number of polynomial variables	n=10	n=20
(a) **Quadratic** {0, 1}.		
Total POPs	100	100
# POPs solved by relaxation order $w=1$	16 (55, 11)	0 (210, 21)
# POPs solved by relaxation order $w=2$	84 (385, 56)	100 (6195, 211)
# POPs solved by $\text{proj}_K([y^1_{e_1}, \dots, y^1_{e_n}])$	75	48
(b) **MAX-CUT**: weights in [0,1] interval.		
Total POPs	100	100
# POPs solved by relaxation order $w=1$	0 (55, 11)	0 (210, 21)
# POPs solved by relaxation order $w=2$	100 (385, 56)	100 (6195, 211)
# POPs solved by $\text{proj}_K([y^1_{e_1}, \dots, y^1_{e_n}])$	28	0
(c) **MAX-CUT**: weights in $[-1, 1]$ interval.		
Total POPs	100	100
# POPs solved by relaxation order $w=1$	5 (55, 11)	0 (210, 21)
# POPs solved by relaxation order $w=2$	95 (385, 56)	100 (6195, 211)
# POPs solved by $\text{proj}_K([y^1_{e_1}, \dots, y^1_{e_n}])$	58	22
(d) **Integer partition**.		
Total POPs	100	100
# POPs solved by relaxation order $w=1$	3 (55, 11)	2 (105, 15)
# POPs solved by relaxation order $w=2$	61 (385, 53)	5 (1470, 106)
# POPs solved by relaxation order $w=3$	36 (847, 176)	21 (6475, 470)
# POPs solved by $\text{proj}_K([y^1_{e_1}, \dots, y^1_{e_n}])$	0	0
# POPs solved by $\text{proj}_K([y^2_{e_1}, \dots, y^2_{e_n}])$	12	0

if y^w is the primal solution of the wth order relaxation, the table also shows for how many of the problems that needed the $(w+1)$th order of the relaxation, the vector $\text{proj}_K([y^w_{e_1}, \dots, y^w_{e_n}])$ was a solution of the original POP. The results indicate for the Quadratic and the MAX-CUT problems, if we are only concerned about the POP solution or bound independent if we found the solution of the relaxation that solves the POP, a good strategy before solving higher order relaxations is to check first the projection of the solution provided by coarser SDP relaxation levels.

The next experiment uses the prolongation operators to provide initial points to SDPT3 to solve the relaxation $w > 1$ for those problems where the first order relaxation does not provide a solution for the original POP. If y^w and X^w are the solutions found by SDPT3 of the primal and dual wth order relaxation respectively for $w > 1$, then we prolongate these coarse solutions using the operators defined in Eqs. (12) and (21), and use these new fine points as initial guesses for SDPT3 to solve the relaxation of order $w+1$. We call this method multilevel algorithm or approach, and we compare it with default SDPT3, i.e., letting SDPT3 calculate the initial points, SeDuMi, and Mosek.

The formulation used by SDPT3 includes an additional primal variable in the relaxation (19) by replacing the constraint $\tilde{M}_w(y) \succeq 0$ by $\tilde{M}_w(y) = S^w$ and $S^w \succeq 0$. Given that SDPT3 is an infeasible interior point method, we need to provide positive definite matrices as starting points, however, the matrices $S^{w+1} = \tilde{M}_w(P^w_y(y^w))$ and $X^{w+1} = \tilde{P}^w_X(X^w)$ are positive semidefinite but not positive definite. We perturb these matrices by using an eigenvalue decomposition and replacing the zero eigenvalues by a small positive number. Preliminary experiments using the prolongated points to solve the fine relaxation showed that even when the prolongated matrices were positive definite, the closer the point was to the boundary of the positive semidefinite cone, the smaller the step sizes calculated by SDPT3 for the initial iterations, making the entire algorithm very slow. We suspect that after the coarse solutions are prolongated, the new feasible points are not close to the central path, which makes the algorithm take extra time getting closer to the central path. This difficulty has been observed in the literature when interior point methods have been combined with warm start strategies, and some approaches to solve this issue has been proposed (see for example Benson & Shanno, 2007 and Skajaa, Andersen, & Ye, 2013). More research is needed in this area to determine the relation between the prolongated points and the central path.[2] We found that getting away of the boundary of the positive semidefinite cone by making all the eigenvalues smaller than 10^{-3} equal to 10^{-3}, was a good trade-off between losing the prolongated points' information and getting larger step sizes in the interior point method. Additionally, for the multilevel method, we changed the early stops of SDPT3 given by the parameter OPTIONS.stoplevel by setting it to zero and we increased the tolerance of the early stop criteria for the infeasibility given in line 721 of the code sqlpmain.m (we replaced 10^{-4} tolerance for 10^{-12}). These changes in the code were done after observing that for some problems the initial step was very small when using the prolongated points, which combined with the small infeasibilities made SDPT3 end prematurely (the same approach is taken in Campos & Parpas, 2018).

Algorithm 1 provides a pseudo-code describing the multilevel method to solve the relaxation of order $w+1$. We define $IPM\left(\{\tilde{A}^{w+1}_{\boldsymbol{\alpha}}, b_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathcal{F}^{w+1}}, y^{w+1}_0, X^{w+1}_0, S^{w+1}_0, \epsilon\right)$ as the function that uses an infeasible interior point method to solve the SDP problem with parameters $\{\tilde{A}^{w+1}_{\boldsymbol{\alpha}}, b_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha} \in \mathcal{F}^{w+1}}$ and initial points $y^{w+1}_0, X^{w+1}_0, S^{w+1}_0$, to a tolerance $\epsilon > 0$.

Algorithm 1 Multilevel method to solve the $(w+1)$th order SDP relaxations (19) and (20) for POPs with {0, 1} or $\{-1, 1\}$ constraints.

Input: Prolongation operators P^w_y and $\tilde{P}^w_X$ defined in Equations (12) and (21), solutions y^w and X^w of the w^{th} order SDP relaxations (19) and (20), parameters $\{\tilde{A}^{w+1}_\alpha, b_\alpha\}_{\alpha \in \mathcal{F}^{w+1}}$, and $\epsilon > 0$.

Procedure:

$y^{w+1}_0 \leftarrow P^w_y(y^w)$
$X^{w+1}_0 \leftarrow \tilde{P}^w_X(X^w)$
$S^{w+1}_0 \leftarrow \tilde{M}_w(y^{w+1}_0)$
$t_X \leftarrow \Lambda(X^{w+1}_0, 10^{-3})$
$t_S \leftarrow \Lambda(S^{w+1}_0, 10^{-3})$
if $t_X > 0$ **then**
  $X^{w+1}_0 \leftarrow \Omega_{X^{w+1}_0}\text{Diag}(10^{-3}, 10^{-3}, \dots, 10^{-3}, \lambda_{t_X+1}(X^{w+1}_0), \dots, \lambda_{\gamma_{w+1}}(X^{w+1}_0))\Omega^\top_{X^{w+1}_0}$
end if
if $t_S > 0$ **then**
  $S^{w+1}_0 \leftarrow \Omega_{S^{w+1}_0}\text{Diag}(10^{-3}, 10^{-3}, \dots, 10^{-3}, \lambda_{t_S+1}(S^{w+1}_0), \dots, \lambda_{\gamma_{w+1}}(S^{w+1}_0))\Omega^\top_{S^{w+1}_0}$
end if
$(y^{w+1}, X^{w+1}) \leftarrow IPM\left(\{\tilde{A}^{w+1}_\alpha, b_\alpha\}_{\alpha \in \mathcal{F}^{w+1}}, y^{w+1}_0, X^{w+1}_0, S^{w+1}_0, \epsilon\right)$

In our case we use SDPT3 as the infeasible IPM, and set the tolerance epsilon equal to 10^{-7}. For those problems where

[2] We also did preliminary experiments using SDPNAL+ (Yang et al., 2015) as our base algorithm. However, the use of the prolongated points did not show any significant improvement. One possible explanation for this result is the fact that although SDPNAL+ and the ADMM algorithm developed in Wen et al. (2010) iterate in the boundary of the semidefinite cone, the points at every iteration satisfy $\langle X^{w+1}, S^{w+1}\rangle = 0$, which is not true for the prolongated points.

Table 2
Average primal-dual infeasibilities, and duality gap for the $(w+1)$th order SDP relaxation using: (1) original prolongated points and (2) the prolongated points after perturbing the eigenvalues of the matrices X^{w+1} and S^{w+1}.

POP/optimality measures	Original		Perturbed		
	d^{w+1}_{infeas}	gap^{w+1}	p^{w+1}_{infeas}	d^{w+1}_{infeas}	gap^{w+1}
Quadratic: $n=10,\ w+1=2$	1.2e−11	2.8e−02	3.7e−02	1.2e−02	6.3e−02
Quadratic: $n=20,\ w+1=2$	1.6e−11	2.9e−02	7.2e−02	1.3e−02	7.1e−02
MAX-CUT ([0,1]): $n=10,\ w+1=2$	7.7e−12	3.0e−01	8.7e−03	4.4e−03	3.5e−01
MAX-CUT ([0,1]): $n=20,\ w+1=2$	1.3e−11	5.2e−01	9.3e−03	4.6e−03	6.1e−01
MAX-CUT ($[-1,1]$): $n=10,\ w+1=2$	1.5e−11	5.0e−02	8.7e−03	4.1e−03	8.0e−02
MAX-CUT ($[-1,1]$): $n=20,\ w+1=2$	1.3e−11	6.0e−02	9.3e−03	4.4e−03	1.2e−01
Integer partition: $n=10,\ w+1=2$	7.8e−09	2.6e−02	8.7e−03	3.3e−03	1.2e−01
Integer partition: $n=10,\ w+1=3$	9.3e−09	4.3e−03	9.3e−03	2.4e−02	2.4e−01
Integer partition: $n=14,\ w+1=2$	3.4e−09	2.7e−02	9.1e−03	3.0e−03	1.8e−01
Integer partition: $n=14,\ w+1=3$	9.3e−10	2.3e−02	9.5e−03	4.0e−02	4.7e−01

the first order relaxation did not find the solution of the POP, we calculated measures for primal infeasibility, dual infeasibility and the duality gap of the prolongated points before and after the eigenvalue perturbation, i.e., we use SDPT3 to find a solution to an accuracy of 10^{-7} for the SDP relaxation of order w and then calculate the optimality measures for the points $(y^{w+1}, S^{w+1}, X^{w+1})$ in Algorithm 1 lines 1–3, and then again but using the matrices calculated in Algorithm 1 lines 7 and 10. Given a point $(y^{w+1}, S^{w+1}, X^{w+1})$ for the $(w+1)$th order SDP relaxations (19) and (20) (not necessarily feasible but satisfying the positive semidefinite constraints), we use the following measures:

- Primal infeasibility:

$$p^{w+1}_{\text{infeas}} := \frac{\|\tilde{M}_{w+1}(y^{w+1}) - S^{w+1}\|}{(1+\gamma^{0.5}_{w+1})}, \tag{22}$$

  where γ_{w+1} is the dimension of the matrix $\tilde{M}_{w+1}(y^{w+1})$.
- Dual infeasibility:

$$d^{w+1}_{\text{infeas}} := \frac{\left(\sum_{\boldsymbol{\alpha}\in\mathcal{F}^{w+1}} \left(\left\langle \tilde{A}^{w+1}_{\boldsymbol{\alpha}}, X^{w+1}\right\rangle - b_{\boldsymbol{\alpha}}\right)^2\right)^{0.5}}{\left(1+\left(\sum_{\boldsymbol{\alpha}\in\mathcal{F}^{w+1}} b^2_{\boldsymbol{\alpha}}\right)^{0.5}\right)}. \tag{23}$$

- Duality gap:

$$\text{gap}^{w+1} := \frac{\left\langle X^{w+1}, S^{w+1}\right\rangle}{\left(1+\sum_{\boldsymbol{\alpha}\in\mathcal{F}^{w+1}} \tilde{b}_{\boldsymbol{\alpha}} y^{w+1}_{\boldsymbol{\alpha}} - [X^{w+1}]_{1,1}\right)}. \tag{24}$$

For a tolerance $\epsilon > 0$, SDPT3 will stop when it has found a point $(y^{w+1}, S^{w+1}, X^{w+1})$ such that $\max\{p^{w+1}_{\text{infeas}}, d^{w+1}_{\text{infeas}}, \text{gap}^{w+1}\} \le \epsilon$.

The average of the optimality measures for every SDP relaxation of order $w > 1$ are shown in Table 2. The primal infeasibility is always zero and therefore we do not report it in the table (Theorem 4.1). As expected, the dual infeasibility using the original prolongation points for the fine relaxation is lower than the 10^{-7} tolerance required for the coarse relaxation (Theorem 4.2 and Lemma 4.1). We observed that the magnitude of the duality gap is of the order of 10^{-2} to 10^{-1}. Although these values are not close to the 10^{-7} accuracy required for our experiments, they are smaller than the observed values achieved by the automatic initial points generated by SDPT3, which for some problems can be of the 10^5 order. When the matrices of the prolongated points are perturbed, the optimality measures for the primal and dual infeasibilities are increased considerably, but they were still lower or at the same levels of the duality gap values. However, as mentioned before, we found that it was necessary to sacrifice these optimality measures to achieve better results when the prolongation points are combined with SDPT3.

Table 3 shows which of the four algorithms solved faster the wth order relaxation for those problems where the $(w-1)$th order relaxation did not providea solution for the POP (the accuracy in every algorithm was set to 10^{-7}). As we are interested in solving the SDP relaxations, we include in these results those problems from where the projection onto the feasible set of the $(w-1)$th order relaxation provided a solution for the POP. In general, Mosek is the fastest algorithm among all the options for large size problems, i.e., problems with more than 10 polynomial variables and/or high order of relaxation. For smaller problems, we can start to see the advantages for the multilevel approach, where for some of the instances, e.g., the Quadratic {0, 1}.

Table 3
Comparison between the times used by SeDuMi, SDPT3, Mosek and Multilevel to solve relaxations of order greater than 1 (accuracy 10^{-7}).

Number of polynomial variables	n=10	n=20
(a) **Quadratic** {0, 1}. Relaxations of order $w=2$.		
Total POPs	84	100
# Solved faster by Mosek	0	100
# Solved faster by SeDuMi	2	0
# Solved faster by SDPT3	0	0
# Solved faster by Multi	82	0
(b) **MAX-CUT**: weights in [0,1] interval. Relaxations of order $w=2$.		
Total POPs	100	100
# Solved faster by Mosek	0	100
# Solved faster by SeDuMi	56	0
# Solved faster by SDPT3	2	0
# Solved faster by Multilevel	42	0
(c) **MAX-CUT**: weights in $[-1,1]$ interval. Relaxations of order $w=2$.		
Total POPs	95	100
# Solved faster by Mosek	53	100
# Solved faster by SeDuMi	13	0
# Solved faster by SDPT3	1	0
# Solved faster by Multilevel	28	0
(d) **Integer partition**. Relaxations of order $w=2$.		
Total POPs	97	26
# Solved faster by Mosek	0	26
# Solved faster by SeDuMi	42	0
# Solved faster by SDPT3	6	0
# Solved faster by Multilevel	49	0
(e) **Integer partition**. Relaxations of order $w=3$.		
Total POPs	36	21
# Solved faster by Mosek	36	21
# Solved faster by SeDuMi	0	0
# Solved faster by SDPT3	0	0
# Solved faster by Multilevel	0	0

The previous experiment shows in absolute terms which algorithm is faster to achieve an accuracy of 10^{-7} (is important to note that the stopping criteria of the algorithms is different, but all the algorithms achieved an accuracy in terms of Eqs. (22)–(24) close to 10^{-7}). However, these results do not show if in general, given a base interior point method, the warm start strategy is useful, e.g., if Mosek is twice as fast as SDPT3 is very unlikely that a warm

Table 4
Time ratios between the Multilevel approach and SDPT3, SeDuMi and Mosek to solve the relaxations of order greater than 1.

Number of polynomial variables	n=10	n=20
(a) **Quadratic** {0, 1}. Relaxations of order $w = 2$.		
Total POPs	84	100
Mean t_{SDPT3}/t_{Multi}	1.68	1.76
Mean t_{Mosek}/t_{Multi}	2.10	0.51
Mean t_{SeDuMi}/t_{Multi}	1.14	16.83
(b) **MAX-CUT**: weights in [0,1] interval. Relaxations of order $w = 2$.		
Total POPs	100	100
Mean t_{SDPT3}/t_{Multi}	1.17	1.04
Mean t_{Mosek}/t_{Multi}	1.27	0.45
Mean t_{SeDuMi}/t_{Multi}	0.96	12.39
(c) **MAX-CUT**: weights in $[-1, 1]$ interval. Relaxations of order $w = 2$.		
Total POPs	95	100
Mean t_{SDPT3}/t_{Multi}	1.27	1.14
Mean t_{Mosek}/t_{Multi}	0.87	0.50
Mean t_{SeDuMi}/t_{Multi}	1.04	13.70
(d) **Integer partition**. Relaxations of order $w = 2$.		
Total POPs	97	26
Mean t_{SDPT3}/t_{Multi}	1.10	1.02
Mean t_{Mosek}/t_{Multi}	1.53	0.52
Mean t_{SeDuMi}/t_{Multi}	1.01	1.96
(e) **Integer partition**. Relaxations of order $w = 3$.		
Total POPs	33	21
Mean t_{SDPT3}/t_{Multi}	1.13	1.08
Mean t_{Mosek}/t_{Multi}	0.27	0.13
Mean t_{SeDuMi}/t_{Multi}	0.54	1.35

start strategy with SDPT3 as the base algorithm will be faster than Mosek. It is more interesting to see relative results between the algorithms, in particular, between SDPT3 and the multilevel approach. Table 4 compares times for the same problems presented in Table 3. Given that SDPT3, SeDuMi and Mosek have different stopping criteria, for this experiment we normalize the criteria to decide when an SDP relaxation is solved. To this end, we use the solution found by SeDuMi, SDPT3 and Mosek, to calculate the primal infeasibility, dual infeasibility and duality gap measures used in SDPT3. Then, for any particular instance, we use the multilevel method to solve three times the SDP relaxation with three different tolerances corresponding to the three other algorithms. We calculate the mean of the ratios of the solution times between each of the three algorithms (SeDuMi, SDPT3 and Multilevel) and the multilevel method. For example, t_{SDPT3}/t_{Multi} for $n = 10$ is equal to 1.68 in Table 4 part (a), which indicates that the mean of the ratio between the time used by SDPT3 and the multilevel method for the 84 problems not solved by the first order relaxation is 1.68. The results give more insight about the speed of the algorithms, of particular interest the relative times between SDPT3 and the multilevel approach. In this case we can see that the multilevel method improves in average the base interior point method for all the problems from a conservative 1.02 for the Integer Partition up to 1.76 for the 20 variables Quadratic problems. The rest of the table shows that Mosek can be up to 10 times faster than the multilevel, nevertheless, this means that in general Mosek can be more than 10 times faster than SDPT3. It would be interesting to use the warm start with Mosek, however, to the best of our knowledge the starting point capability is not offered by this solver.

6. Conclusions

Using SDP relaxations for polynomial optimization problems has been proved to be a powerful tool to solve other-wise hard non-convex problems. This paper proposes a new approach to exploit the usually unused information contained in the lower levels of the Lasserre hierarchy. The new prolongation operators relating the lower and higher levels are simple and easy to implement, and our numerical experiments show that they can be useful as an approximate solution by themselves, or as an initial point to be used along with an interior point method. When the latter version is implemented, we do not claim that the warm start method (which we have referred as the multilevel approach) is better than any other IPM or any other alternative to solve POPs (like the algorithm Biq Mac for MAX-CUT problems developed in Rendl, Rinaldi, & Wiegele, 2010). In fact, depending on the particular POP that we are trying to solve and its size, other solvers can perform better than SDPT3 and our multilevel version. However, we can improve the efficiency of the underlying IPM, in our case SDPT3, and given the inexpensive cost of calculating the prolongation points once the solution of a lower relaxation has been found, it is worth to use them as initial points when no more information is available. Recently, new and promising SDP relaxations have been proposed (Lasserre et al., 2017). These new hierarchies have shown good numerical results when compared with the classical Lasserre hierarchy, and therefore it will be interesting to implement the multilevel ideas in that framework.

Acknowledgments

This work was funded by Engineering & Physical Sciences Research Council grant numbers EP/P016871/1 and EP/M028240/1.

References

Ahmadi, A. A., & Majumdar, A. (2017). DSOS and SDSOS optimization: More tractable alternatives to sum of squares and semidefinite optimization. arXiv:1706.02586.
Benson, H. Y., & Shanno, D. F. (2007). An exact primal–dual penalty method approach to warmstarting interior-point methods for linear programming. *Computational Optimization and Applications, 38*(3), 371–399.
Boukouvala, F., Misener, R., & Floudas, C. A. (2016). Global optimization advances in mixed-integer nonlinear programming, MINLP, and constrained derivative-free optimization, CDFO. *European Journal of Operational Research, 252*(3), 701–727.
Burer, S., & Monteiro, R. (2003). A nonlinear programming algorithm for solving semidefinite programming via low-rank factorization. *Mathematical Programming (series B), 95*, 329–357.
Campos, J. S., & Parpas, P. (2018). A multigrid approach to SDP relaxations of sparse polynomial optimization problems. *SIAM Journal on Optimization, 28*(1), 1–29.
Caprara, A. (2008). Constrained 0-1 quadratic programming: Basic approaches and extensions. *European Journal of Operational Research, 187*(3), 1494–1503.
van Dam, E. R., & Sotirov, R. (2015). Semidefinite programming and eigenvalue bounds for the graph partition problem. *Mathematical Programming, 151*(2), 379–404.
de Klerk, E. (2010). Exploiting special structure in semidefinite programming: A survey of theory and applications. *European Journal of Operational Research, 201*(1), 1–10.
De Klerk, E., & Laurent, M. (2011). On the lasserre hierarchy of semidefinite programming relaxations of convex polynomial optimization problems. *SIAM Journal on Optimization, 21*(3), 824–832.
Lasserre, J. B. (2001). Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization, 11*(3), 796–817.
Lasserre, J. B. (2002). An explicit equivalent positive semidefinite program for nonlinear 0-1 programs. *SIAM Journal on Optimization, 12*(3), 756–769.
Lasserre, J. B. (2006). Convergent SDP-relaxations in polynomial optimization with sparsity. *SIAM Journal on Optimization, 17*(3), 822–843.
Lasserre, J. B. (2009). Convexity in semialgebraic geometry and polynomial optimization. *SIAM Journal on Optimization, 19*(4), 1995–2014.
Lasserre, J. B. (2016). A MAX-CUT formulation of 0/1 programs. *Operations Research Letters, 44*(2), 158–164.
Lasserre, J. B., Toh, K.-C., & Yang, S. (2017). A bounded degree SOS hierarchy for polynomial optimization. *EURO Journal on Computational Optimization, 5*(1–2), 87–117.
Nie, J. (2014). Optimality conditions and finite convergence of lasserres hierarchy. *Mathematical programming, 146*(1–2), 97–121.
Rendl, F., Rinaldi, G., & Wiegele, A. (2010). Solving max-cut to optimality by intersecting semidefinite and polyhedral relaxations. *Mathematical Programming, 121*(2), 307.
Schweighofer, M. (2005). Optimization of polynomials on compact semialgebraic sets. *SIAM Journal on Optimization, 15*(3), 805–825.
Skajaa, A., Andersen, E. D., & Ye, Y. (2013). Warmstarting the homogeneous and self--dual interior point method for linear and conic quadratic problems. *Mathematical Programming Computation, 5*(1), 1–25.
Sturm, J. F. (1999). Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software, 11*(1–4), 625–653.

Toh, K.-C., Todd, M. J., & Tütüncü, R. H. (2012). On the implementation and usage of SDPT3–a MATLAB software package for semidefinite-quadratic-linear programming, version 4.0. In *Handbook on semidefinite, conic and polynomial optimization* (pp. 715–754). Springer.

Waki, H., Kim, S., Kojima, M., & Muramatsu, M. (2006). Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. *SIAM Journal on Optimization, 17*(1), 218–242.

Waki, H., Kim, S., Kojima, M., Muramatsu, M., & Sugimoto, H. (2008). Algorithm 883: SparsePOP—A sparse semidefinite programming relaxation of polynomial optimization problems. *ACM Transactions on Mathematical Software, 35*(2), 15.

Weisser, T., Lasserre, J. B., & Toh, K.-C. (2017). Sparse-BSOS: A bounded degree SOS hierarchy for large scale polynomial optimization with sparsity. *Mathematical Programming Computation*. doi:10.1007/s12532-017-0121-6.

Wen, Z., Goldfarb, D., & Yin, W. (2010). Alternating direction augmented Lagrangian methods for semidefinite programming. *Mathematical Programming Computation, 2*, 203–230.

Yang, L., Sun, D., & Toh, K.-C. (2015). SDPNAL+: A majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. *Mathematical Programming Computation, 7*(3), 331–366.